Phenotypic Mutation Rates and the Abundance of Abnormal Proteins in Yeast

Martin Willensdorfer[1,2*], Reinhard Bürger[1,3], Martin A. Nowak[1,4]

1 Program for Evolutionary Dynamics, Harvard University, Cambridge, Massachusetts, United States of America, **2** Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts, United States of America, **3** Department of Mathematics, University of Vienna, Vienna, Austria, **4** Department of Mathematics, Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts, United States of America

Phenotypic mutations are errors that occur during protein synthesis. These errors lead to amino acid substitutions that give rise to abnormal proteins. Experiments suggest that such errors are quite common. We present a model to study the effect of phenotypic mutation rates on the amount of abnormal proteins in a cell. In our model, genes are regulated to synthesize a certain number of functional proteins. During this process, depending on the phenotypic mutation rate, abnormal proteins are generated. We use data on protein length and abundance in *Saccharomyces cerevisiae* to parametrize our model. We calculate that for small phenotypic mutation rates most abnormal proteins originate from highly expressed genes that are on average nearly twice as large as the average yeast protein. For phenotypic mutation rates much above 5×10^{-4}, the error-free synthesis of large proteins is nearly impossible and lowly expressed, very large proteins contribute more and more to the amount of abnormal proteins in a cell. This fact leads to a steep increase of the amount of abnormal proteins for phenotypic mutation rates above 5×10^{-4}. Simulations show that this property leads to an upper limit for the phenotypic mutation rate of approximately 2×10^{-3} even if the costs for abnormal proteins are extremely low. We also consider the adaptation of individual proteins. Individual genes/proteins can decrease their phenotypic mutation rate by using preferred codons or by increasing their robustness against amino acid substitutions. We discuss the similarities and differences between the two mechanisms and show that they can only slow down but not prevent the rapid increase of the amount of abnormal proteins. Our work allows us to estimate the phenotypic mutation rate based on data on the fraction of abnormal proteins. For *S. cerevisiae*, we predict that the value for the phenotypic mutation rate is between 2×10^{-4} and 6×10^{-4}.

Citation: Willensdorfer M, Bürger R, Nowak MA (2007) Phenotypic mutation rates and the abundance of abnormal proteins in yeast. PLoS Comput Biol 3(11): e203. doi:10.1371/journal.pcbi.0030203

Introduction

Every biological organism is built according to information stored in its genome. Genomes composed of billions of base pairs are not unusual. This information has to be duplicated during cell replication. Since replication errors can have devastating effects, DNA replication needs to be very accurate. Estimates for error rates in Eukaryotes are as low as 5×10^{-4} errors per base pair per replication [1]. But even flawless genetic information is useless if the cell is not able to synthesize functional proteins. Transcription and translation, the two processes involved in decoding DNA, have to be sufficiently accurate to allow a cell to build a reliable protein machinery. We refer to errors that occur during transcription and translation as phenotypic errors, and to errors that occur during DNA replication as genotypic errors. Most phenotypic errors are introduced during translation when ribosomes translate RNA sequences into amino acid sequences [2,3]. The accuracy of translation depends on the considered codon and context. In *Escherichia coli* it can range from 5×10^{-4} to 1×10^{-4} (see Table 1 for some examples), with 5×10^{-4} as a commonly used estimate for the average frequency of errors per codon [4,5]. In comparison, Blank et al. [2] measured an *E. coli* error rate during transcription of 5×10^{-6}.

Measuring the genotypic mutation rate is easier than measuring the phenotypic mutation rate. Estimates of genotypic mutation rates exist for many organisms. The data show that the number of mutations per genome per replication is constant for a wide range of organisms [1]. This is in agreement with theoretical results that suggest that the number of errors per replication per genome have to be below a certain error threshold to avoid an error catastrophe at which the propagation of genetic information becomes impossible [6–9]. There are many theoretical approaches for studying the evolution of genotypic mutation rates [10–14], and one can tentatively claim that we have a basic understanding of what governs the evolution of genotypic mutation rates.

This is not the case for phenotypic mutation rates. Apparently, very little theoretical work has been devoted to this topic. A notable exception is Wilke and Drummond [15], who study translational robustness and the evolution of gene-specific phenotypic mutation rates. Their work predicts selection for proteins that fold properly despite mistranslation and provides an explanation for the fact that highly expressed genes evolve slower. A closely related study, and the starting point for this investigation, is Bürger et al. [16]. They studied a model in which the total number of synthesis

Editor: Lauren Ancel Meyers, University of Texas Austin, United States of America

Received January 24, 2007; **Accepted** September 5, 2007; **Published** November 23, 2007

A previous version of this article appeared as an Early Online Release on September 6, 2007 (doi:10.1371/journal.pcbi.0030203.eor).

Copyright: © 2007 Willensdorfer et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abbreviations: LL, log-likelihood

* To whom correspondence should be addressed. E-mail: ma.wi@gmx.at

Author Summary

A functional protein machinery, built from genetic information, is central to every living organism. Surprisingly, the decoding of genes into amino acid sequences is fairly inaccurate. Errors in this process (phenotypic mutations) are several orders of magnitude more frequent than errors during DNA replication (genotypic mutations). Many researchers have explored the evolution of genotypic mutation rates, but there are as yet few investigations into the evolutionary dynamics of phenotypic mutation rates. Here we present a mathematical model that describes the effect of phenotypic mutation on the amount of abnormal proteins in cells. We parameterize our model using data from yeast (*Saccharomyces cerevisiae*). We show that for phenotypic mutation rates above 5×10^{-4} per amino acid, the error-free synthesis of large proteins becomes nearly impossible. We estimate the phenotypic mutation rate of *S. cerevisiae* to be between 2×10^{-4} and 6×10^{-4} per amino acid.

attempts to produce sufficiently many functional proteins is limited and showed that the selection pressure to reduce the phenotypic mutation rate below a certain threshold vanishes.

In addition, empirical information is scarce. Most measurements of phenotypic mutation rates are limited to *E. coli* [5]. For fast-growing *E. coli* laboratory strains, a correlation was found between ribosomal accuracy and ribosomal kinetics [3,17]. This suggests that the (high) phenotypic mutation rates are a result of a cost–benefit tradeoff. More accurate ribosomes reduce speed of translation and are hence disadvantageous. Natural isolates, however, do not show such a correlation between ribosomal accuracy and ribosomal kinetics. They display a wide diversity of ribosomal kinetic properties and growth rates which suggests that the tradeoff between accuracy and kinetics is not limiting in natural populations [17–19]. Apparently, natural populations are not so obsessed with optimizing translation kinetics for fast growth under laboratory conditions. This is not surprising considering that the estimated doubling time of, for example, intestinal *E. coli* (40 h) is substantially longer than the doubling time of laboratory strains (0.5 h) [20].

Hence, it is not clear if the optimization of translation kinetics is governing the evolution of phenotypic mutation rates. In this paper we analyse phenotypic mutations from a genomic/proteomic viewpoint. In particular, we derive and analyze a model that allows us to calculate the amount of abnormal proteins in a cell as a function of the phenotypic mutation rate. We also evolve genotypic and phenotypic mutation rates in computer simulations that are based on properties of the *Saccharomyces cerevisiae* genome/proteome. We discover that the current estimate for global phenotypic mutation rates of 5×10^{-4} is at a value where the amount of erroneous proteins begins to increase exponentially with the mutation rate. Further, at this value we observe a change in the kind of genes that contribute the most to erroneous proteins. For phenotypic mutation rates below 5×10^{-4}, erroneous proteins from highly expressed genes are frequent. Above 5×10^{-4}, however, erroneous proteins from large genes begin to dominate. Finally, we study models in which individual proteins can decrease their phenotypic mutation rate by using preferred codons or evolve robustness against amino acid substitutions. We point out the similarities and differences between the two mechanisms and show how an increase of the amino acid substitution rate above 5×10^{-4} affects the adaptation of highly expressed proteins.

Table 1. Amino Acid Substitution Rates During Protein Synthesis in *E. coli*

Codon	Amino Acid		Frequency	Reference
	Normal	Substituted		
AAC	Asn	Lys	4×10^{-4}	[44]
AAU	Asn	Lys	2×10^{-3}	[44]
CGU	Arg	Cys	1×10^{-3}	[45]
CGU/C	Arg	Cys	5×10^{-5}	[46]
GGC	Gly	Ser	1×10^{-3}	[47]
UGG	Trp	Cys	3×10^{-3}	[45]
			4.5×10^{-4}	[4]

See Table 1 of Parker [5] for more information.
doi:10.1371/journal.pcbi.0030203.t001

Materials and Methods

In the following, we develop and analyze a model regarding the evolution of phenotypic mutation rates. We use data from *S. cerevisiae* to parameterize our model and calculate here relevant properties of the available yeast data.

The genotypic mutation rate in *S. cerevisiae* is approximately 2.2×10^{-10} mutations per base pair per replication [1]. Our model requires the number of deleterious mutations per codon per replication as the unit for the genotypic mutation rate. Since each codon is composed of three nucleotide acids and $438/576 \approx 3/4$ single site mutations are nonsynonymous [21], the mutation rate per codon is given by $3 \times 3/4 \times 2.2 \times 10^{-10} = 4.95 \times 10^{-10}$. Of these nonsynonymous mutations, about 10% to 60% [22] are deleterious, placing the genotypic mutation rate somewhere between 4.95×10^{-11}, and 2.97×10^{-10} deleterious mutations per codon per replication. For simplicity, we use 1×10^{-10}. The phenotypic mutation rate in yeast appears to be similar to the mutation rates measured in *E. coli* [23]. The mutation rate is therefore likely to range from 1×10^{-5} to 5×10^{-3}, with 5×10^{-4} as an estimate for the global phenotypic mutation rate [5].

To parameterize our model, we need the length n_i and abundance y_i of each protein in yeast. Complete genomic sequences [24] provide the length, n_i, of each protein of an organism. We only consider reading frames from the *Saccharomyces* Genome Database [24] that have been classified as nonspurious by Ghaemmaghami et al. [25]. This leaves us with 5,675 open reading frames (proteins) that have an average length of 496 amino acids. The effective genome is $n = \sum_i n_i = 2.81 \times 10^6$ residues long. Information about the abundance, y_i, of proteins is provided by Ghaemmaghami et al. [25]. From their data we calculate that the total amount (measured in number of amino acids) of functional proteins is given by $y = \sum_i y_i n_i = 2.03 \times 10^{10}$ and that $y = \sum_i y_i n_i^2 = 1.34 \times 10^{13}$ (this expression is relevant for Equation 8).

Results

The Model

We consider a large population of asexual single-cell organisms (cells, for short), each with one DNA chromosome

and K genes of (possibly) different length and expression level. Gene i ($i = 1,\ldots,K$) is n_i amino acids long. During one cell cycle, y_i, functional proteins have to be synthesized from gene i. We assume that regulation of gene expression guarantees that the gene is expressed until y_i functional proteins are present. A cell with error-free transcription and translation will therefore synthesize exactly y_i proteins of gene i. Since we are interested in variation within a population, we can normalize the fitness of such a cell to 1 and consider only relative fitnesses. Any cost of synthesizing all the functional proteins is accounted for in the fitness value of 1. Cells, however, have a phenotypic mutation rate $u > 0$, which denotes the probability (per codon) that protein synthesis is erroneous and produces a nonfunctional protein. Hence, a phenotypic mutation is a deleterious amino acid substitution that occurs during protein synthesis. We assume that phenotypic mutations are independent of each other. Hence, the probability of synthesizing a nonfunctional protein from gene i is given by $u_i = 1 - (1-u)^{n_i}$. Let x_i denote the number of nonfunctional proteins that have been synthesized until y_i functional proteins were made. Then a cell synthesizes $\sum_i y_i$ functional and $\sum_i x_i$ nonfunctional proteins. It uses $y = \sum_i n_i y_i$ amino acids to synthesize functional proteins and $x = \sum_i n_i x_i$ amino acids to synthesize nonfunctional proteins.

We are interested in the cost of erroneous protein synthesis in natural populations. Natural populations grow much slower than laboratory cultures. The growth of bacteria is limited by the rate of protein synthesis per ribosome. In slow-growing bacteria, the rate of protein synthesis per ribosome is, because of limiting amounts of charged tRNAs, almost 50% lower than in fast-growing bacteria [26]. We will therefore base our cost function on the effects of the phenotypic mutation rate on the availability of charged tRNAs.

Most erroneous proteins are identified as abnormal and are degraded rapidly [27]. In this case, the amino acids that have been used to synthesize the erroneous proteins can be recycled to charge new tRNAs. This constant turnover, however, will diminish the pool of charged tRNAs by an amount depending on x. Hence, the cost of phenotypic mutations is a function of x. We will use $\eta(x)$ to denote the cost of erroneous protein synthesis. Even though we have motivated $\eta(x)$ by the cost of depleting the tRNA pool, it can account for other possible costs of erroneous proteins synthesis as well. Examples include toxic effects of aggregates of misfolded proteins, the waste of metabolic energy (ATP/GTP), or usage of the ribosomal machinery to synthesize nonfunctional instead of functional proteins. Overall, the fitness of a cell that uses x amino acids to synthesize nonfunctional proteins is given by $1 - \eta(x)$. It is unnecessary to explicitly consider the cost of protein synthesis of functional proteins, c_y, because y is constant and, hence, $1 - c_y - \tilde{\eta}(x) \propto 1 - \frac{\tilde{\eta}(x)}{1-c_y} = 1 - \eta(x)$, where $\tilde{\eta}(x)$ denotes the costs if the costs of synthesis of functional proteins are not included.

Besides phenotypic, we also take genotypic mutations into account. This allows us to directly compare the cost and evolution of phenotypic and genotypic mutation rates. Genotypic mutations introduce deleterious mutations at rate μ per codon, that is, gene i replicates successfully with probability $(1-\mu)^{n_i}$, and all genes are replicated successfully with probability $(1-\mu)^n$, where $n = \sum_i n_i$ can be interpreted as the effective genome size, i.e., the total number of amino acids encoded by all K genes. We assume that the population is large enough so that deleterious mutations cannot spread. We ignore mutations that recover the wild type.

In the following, the average fitness of a population with phenotypic and genotypic mutation rates u and μ per codon will play a central role. In the absence of genotypic mutations, the mean fitness is $1 - \bar{\eta}(u)$, where $\bar{\eta}(u) = \sum_x \eta(x) p_u(x)$, and $p_u(x)$ denotes the probability that a cell with a phenotypic mutation rate u uses x amino acids to synthesize abnormal proteins. Genotypic mutations reduce this mean fitness. Because we ignore back mutations, at mutation–selection balance the mean fitness is reduced by the factor $(1-\mu)^n$, and this factor is independent of the actual fitness of the mutants. This is a special case of Haldane's principle for the mutation load (see the section Mutation Rates at Equilibrium). The mean fitness at mutation–selection balance is given by

$$\bar{f}(\mu, u) = (1-\mu)^n [1 - \bar{\eta}(u)]. \quad (1)$$

As a consequence, we get the following formula for the cost C_u of phenotypic mutations, which is defined as the difference between the expected mean fitness of a population with and without phenotypic mutations:

$$C_u = \bar{f}(\mu, 0) - \bar{f}(\mu, u) = (1-\mu)^n \bar{\eta}(u). \quad (2)$$

Similarly, the cost C_μ of genotypic mutations is

$$C_\mu = \bar{f}(0, \mu) - \bar{f}(\mu, u) = [1 - (1-\mu)^n][1 - \bar{\eta}(u)]. \quad (3)$$

Evolving Mutation Rates

We allow the mutation rates μ and u to evolve. Our aim is to derive approximations for the mutation rates that evolve in the long run. For this purpose, it is convenient to consider μ and u as quantitative traits with values between 0 and 1. Both traits have the potential to evolve due to new mutations which are assumed to occur at constant rates π_μ and π_u, respectively. These mutations will primarily increase the genotypic and phenotypic mutation rates, but some will decrease them. Because the precise form of these mutation distributions does not enter our approximate formulas given here, we introduce them only in the context of our computer simulation model below. Mutation is counteracted by selection because, on average, cells with a higher mutation rate have a lower fitness. Applying Haldane's principle, it can be shown (see the section Mutation Rates at Equilibrium) that the evolved genotypic mutation rate, $\hat{\mu}$, can be approximated as

$$\hat{\mu} \approx \pi_\mu / n. \quad (4)$$

Similarly, the phenotypic mutation rate equilibrates to a value, $\hat{u}$, which is obtained (approximately) by solving the equation

$$\pi_u = \bar{\eta}(u) \quad (5)$$

for u. If, as is likely the case, $\bar{\eta}$ is monotone increasing, this solution $\hat{u}$ is uniquely determined. By taking the ratio of Equation 4 and Equation 5, and performing a simple rearrangement, we obtain

$$\hat{\mu} n \approx \bar{\eta}(\hat{u}) \frac{\pi_\mu}{\pi_u}. \quad (6)$$

The term $\hat{\mu} n$ gives the number of mutations per genome per replication and is surprisingly constant for a wide range of organisms [1]. The rates π_μ and π_u at which the genotypic and phenotypic mutation rates are changed will primarily

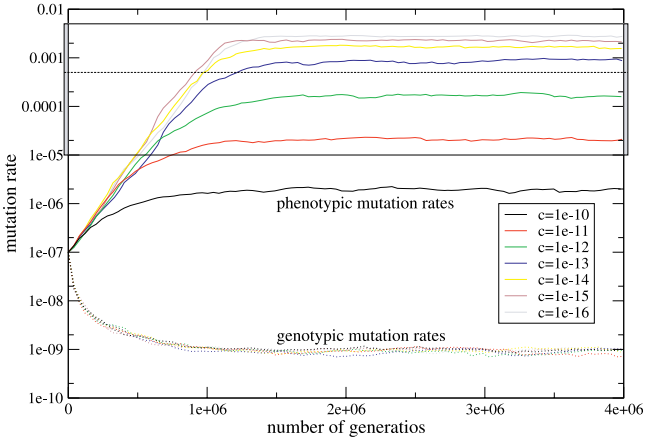

Figure 1. Evolving Phenotypic and Genotypic Mutation Rates

The evolution of phenotypic (solid lines) and genotypic (dotted lines) mutation rates is shown. For each of seven different values for the cost of erroneous proteins, *c*, we conducted ten simulation runs and calculated an average evolutionary trajectory (see text for more details). As expected, only the phenotypic mutation rate is affected by changes in *c*. Near $u = 2 \times 10^{-3}$ we observe an upper limit for the phenotypic mutation rate. Even large changes in *c* affect the phenotypic mutation rate only marginally (compare grey and brown lines). This (effective) upper bound for *u* is the result of a rapid, nonlinear increase in abnormal proteins as a function of *u* (see Figure 2). The (grey) box indicates the possible range of phenotypic mutation rates (1×10^{-5} – 5×10^{-3}, according to Parker [5]) with 5×10^{-4} (dashed line) as a commonly used estimate for the global error rate.

doi:10.1371/journal.pcbi.0030203.g001

depend on the number of genes (and their length) involved in DNA replication and protein synthesis, respectively. In our model, for a given set of parameters, genotypic and phenotypic mutation rates evolve independently. Although we focus on the evolution of phenotypic mutation rates, for the analysis of our simulations it proved useful to also keep track of the genotypic mutation rate. Its equilibrium value is independent of the cost function η and can be used to estimate the effectiveness of selection (i.e., the drift–selection equilibrium value of $\bar{f}$) for given population size and given values of π_u and π_μ .

A more detailed expression for the evolved phenotypic mutation rate can be obtained by approximating the cost function $\eta(x)$, presumably concave, by a linear one. Thus, let us assume henceforth $\eta(x) = cx$, so that fitness decreases linearly with the number of amino acids used to synthesize erroneous proteins. Here, *c* measures the costs per codon to synthesize an abnormal protein. As a consequence, we can write $\bar{\eta}(x) = c\bar{x}$. In the Discussion we will address how nonlinear cost functions might affect our results.

Since, according to our model, a cell produces proteins until exactly y_i functional proteins of gene *i* are synthesized, the number x_i of nonfunctional proteins produced during the process follows the negative binomial distribution $NB(y_i, 1 - u_i)$. Here, $1 - u_i$ gives the probability of a successful protein synthesis. The expected value of x_i is $\bar{x}_i = y_i u_i/(1 - u_i)$. Hence, $\bar{x}$, the expected value of *x*, is given by

$$\bar{x} = \sum_i n_i \bar{x}_i = \sum_i n_i y_i [1 - (1-u)^{n_i}]/(1-u)^{n_i} = \sum_i n_i y_i [(1-u)^{-n_i} - 1]. \quad (7)$$

For $un_i < 1$, we obtain

$$\bar{x} \approx u \sum_i n_i^2 y_i, \quad (8)$$

and $\bar{\eta}(u) \approx uc \sum_i n_i^2 y_i$. Therefore, Equation 6 can be rearranged to

$$\frac{\hat{u}}{\hat{\mu}} \approx \frac{\sum_i n_i}{c \sum_i n_i^2 y_i} \frac{\pi_u}{\pi_\mu}. \quad (9)$$

This illustrates nicely the factors that determine the ratio of the evolved genotypic and phenotypic mutation rates. This ratio depends on (i) the effective genome size, $\sum_i n_i$, (ii) the average total cost of abnormal protein synthesis, $c \sum n_i^2 y_i$, and (ii) the ratio at which mutations of the two mutation rates occur. Below, we will use computer simulations to complement these analytical considerations.

Simulating the Evolution of Mutation Rates

We simulate the evolution of phenotypic and genotypic mutation rates based on the model introduced above. The main difference is that we use a finite (effective) population of size $N = 10^4$ and that we specify our mutation distributions for the mutation rates. Each generation, *N* organisms are selected (with replacement) for reproduction with probabilities proportional to their fitness (Wright-Fisher model of drift and selection). Fitness is calculated as $1 - \bar{\eta}(u) = 1 - c\bar{x}$, where $\bar{x}$ gives the expected amount (in amino acids) of abnormal proteins. Protein length and abundances are taken from *S. cerevisiae* (see Materials and Methods). The number of expected abnormal proteins is calculated according to Equation 7. To avoid the fixation of genotypic mutants for high genotypic mutations rates at the beginning of the simulations, we set the fitness of genotypic mutants to zero. As predicted by the theory, the fitness of the genotypic mutants does not affect the equilibrium mutation rates (see the section Effect of Initial Values and Parameters on the Simulation Results).

The initial population is homogeneous with equal phenotypic and genotypic mutation rate. To allow the evolution of mutation rates, we change (mutate) *u* and μ with probabilities π_u and π_μ, respectively. Unless otherwise mentioned, we assume $\pi_\mu = \pi_u = 10^{-4}$. In the section Effect of Initial Values and Parameters on the Simulation Results, we show that changes in the initial values of *u* and μ do not affect the results of our simulations and changes in π_u and π_μ affect them as predicted by the theory. Since beneficial mutations, that is, mutations decreasing the mutation rate, are generally rare, we increase the mutation rate with a probability of 0.99 (conditional on a mutation event). In this case, we draw a new phenotypic mutation rate from a beta distribution $B(a = 1, b = 9)$ on $[u, 2u]$ (or on $[\mu, 2\mu]$ for genotypic mutation rates). Hence, the average increase is 10% and small changes are more frequent than large ones. Similarly, in case of a decrease, we draw the new mutation rate from a reflected beta distribution $B(a = 1, b = 9)$ on $[0, u]$ (or $[0, \mu]$). Hence, small changes are again more likely than large ones.

During a simulation run, we kept track of the ancestry of each individual. After 4×10^6 generations, we calculate the most recent common ancestor of the population and determine its line of descent. We can use this line of descent to observe the evolution of phenotypic and genotypic mutation rates. For each parameter combination, we con-

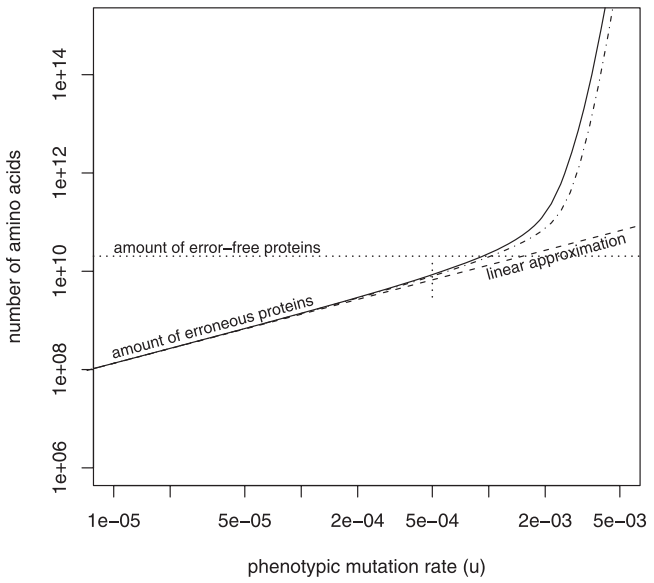

Figure 2. The Amount of Erroneous Proteins as a Function of the Phenotypic Mutation Rate, u

The solid curve shows the expected number $\bar{x}$ of amino acids required to synthesize abnormal proteins according to Equation 7 with values for n_i and y_i from yeast (see Methods and Materials). The dash-dotted curve shows the total number $\sum_i \bar{x}_i$ of abnormal proteins. For better comparison, we scaled the number of proteins so that the dash-dotted and solid curves meet at $u = 10^{-5}$. The dashed line shows the linear approximation to $\bar{x}$ (see Equation 8). The dotted line indicates the amount (in amino acids) of functional proteins in a yeast cell, which equals 2.029×10^{10}. Near $u = 5 \times 10^{-4}$ (the estimate for the global phenotypic mutation rate), the linear approximation begins to deviate noticeably from the exact value. A doubling of u at this point would result in more erroneous than error-free proteins. Another doubling would result in more than seven times as many erroneous than error-free proteins. This nonlinear increase is also observed if one considers the number of abnormal proteins (dash-dotted curve).

doi:10.1371/journal.pcbi.0030203.g002

ducted ten runs that differ only with respect to the seed for the random number generator. Based on the ten trajectories, we compute an expected evolutionary trajectory for each parameter combination by calculating the geometric average of μ and u.

Figure 1 shows these average trajectories of u and μ for seven sets of simulations that use different values for c, ranging from 1×10^{-16} to 1×10^{-10}, with increments of one order of magnitude. For all seven cost values, the genotypic mutation rates (lower lines) show essentially identical behavior. They decrease from the initial value of 1×10^{-7} to about 1.01×10^{-9} which leaves the cost of genotypic mutations at about $C_\mu = 2.84 \times 10^{-3}$.

The equilibrium value of the phenotypic mutation rate depends strongly on the cost of abnormal proteins. For the given values of $n = 2.81 \times 10^6$ and $\sum_i y_i n_i^2 = 1.34 \times 10^{13}$ (see Materials and Methods) and because we assume $\pi_\mu = \pi_u$, Equation 9 predicts $\hat{u}/\hat{\mu} = 2.10 \times 10^{-7}/c$. For $c = 10^{-10}$, this equals 2.10×10^3 and is indeed very close to the observed value of $\hat{u}/\hat{\mu} = 2.00 \times 10^{-6}/1.01 \times 10^{-9} = 1.97 \times 10^3$ (see black curves in Figure 1). For this set of simulations, the phenotypic mutation rate evolves to a level at which the cost of phenotypic mutations is $C_u = 2.67 \times 10^{-3}$, which is very close to C_μ.

As expected, a decrease in the cost of abnormal proteins, c, results in an increase of the phenotypic mutation rate. But, apparently, there is an upper limit for phenotypic mutation rates above which a further decrease of c does not increase $\hat{\mu}$ much more (compare brown and grey curves in Figure 1). Also, for very small c, Equation 9 becomes inaccurate. For example, if $c = 10^{-13}$, we get $\hat{u}/\hat{\mu} = 8.74 \times 10^{-4}/1.01 \times 10^{-9} = 0.87 \times 10^6$ instead of $\hat{u}/\hat{\mu} = 2.11 \times 10^6$.

We can explain both observations by considering the average number $\bar{x}$ of amino acids required to synthesize nonfunctional proteins (Equation 7) and the total number $\sum_i \bar{x}_i$ of nonfunctional proteins. For brevity and in distinction to the number of nonfunctional proteins, $\sum_i \bar{x}_i$, we henceforth refer to $\bar{x} = \sum_i n_i \bar{x}_i$ as the *amount* of abnormal proteins. (We use the number of amino acids as the unit for the amount of proteins.) Figure 2 displays these quantities as functions of u (solid line for $\bar{x}$, dash-dotted line for $\sum_i \bar{x}_i$) as well as the linear approximation (Equation 8) to $\bar{x}$ (dashed line). As one can see, the linear approximation becomes inaccurate if $u > 5 \times 10^{-4}$, and $\bar{x}$ begins to grow exponentially with u. Consequently, even a small increase in u will cause a tremendous increase in $\bar{x}$. Even if abnormal proteins are not very costly, the rapid increase in $\bar{x}$ prevents a further increase of u.

We used $(1-u)^{-n_i} - 1 \approx u n_i$ to linearize $\bar{x}$. This approximation is only accurate if $u n_i$ is sufficiently small. For a protein length of about 890 amino acids (see below for an explanation why we chose 890) and a phenotypic mutation rate of 5×10^{-4}, we have $u n_i = 0.445$, which is apparently too large for the approximation to be accurate. For $n_i = 890$ and $u = 5 \times 10^{-4}$, the exact value for $(1-u)^{-n_i} - 1$ is 0.56; this illustrates the difference between the true value of $\bar{x}$ and its linear approximation at $u = 5 \times 10^{-4}$. It is important to emphasize that the observed upper bound for u is a consequence of the protein length distribution and the expression profile of the organism, as we will see below.

Components of $\bar{x}$

It would be interesting to know the length and expression level of the genes that contribute most to the amount of abnormal proteins in a cell. For a given phenotypic mutation rate, it is easy to calculate $n_i \bar{x}_i$, the amount (in amino acids) of erroneous proteins that were produced from gene i. To determine which gene lengths and expression levels are most important for $\bar{x}$, we calculate weighted averages of n_i and y_i. As weights, we use the amount of erroneous proteins that stem from gene i, that is, $n_i \bar{x}_i$. Hence, we have

$$\sum_i n_i n_i \bar{x}_i / \sum_i n_i \bar{x}_i \tag{10}$$

and

$$\sum_i y_i n_i \bar{x}_i / \sum_i n_i \bar{x}_i \tag{11}$$

as indicators of the average protein length and expression level, respectively, that are most important for the amount of abnormal proteins in the cell. Since the expected number of abnormal proteins, $\bar{x}_i$, depends on the phenotypic mutation rate, the weighted averages are functions of u as well. How they change as a function of u is shown in Figure 3. Interestingly, the weighted average of n_i for very small u is about 890 and nearly twice as large as 496, the average protein length in yeast. Even more interestingly, this value

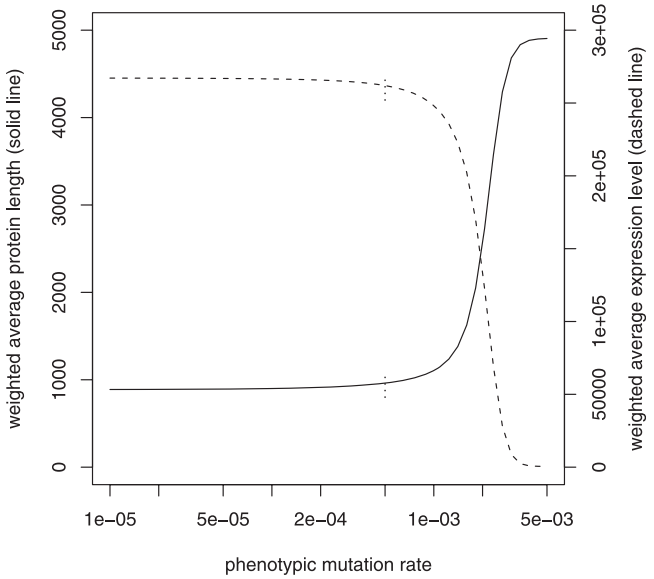

Figure 3. Weighted Averages of Protein Length (10) and Expression Level (11)

To determine which protein and expression levels are most relevant for the amount of abnormal proteins in a cell, we calculated weighted averages of protein length (solid line) and expression level (dashed line). As weights, we used the amount of expected abnormal proteins, $n_i\bar{x}_i$. For small phenotypic mutation rates, highly expressed proteins are most relevant for the amount of abnormal proteins in a cell. This begins to change at $u = 5 \times 10^{-4}$, when lowly expressed, large proteins begin to dominate $\bar{x}$. Inaccurate protein synthesis makes it practically impossible to synthesize these larger proteins error-free.

doi:10.1371/journal.pcbi.0030203.g003

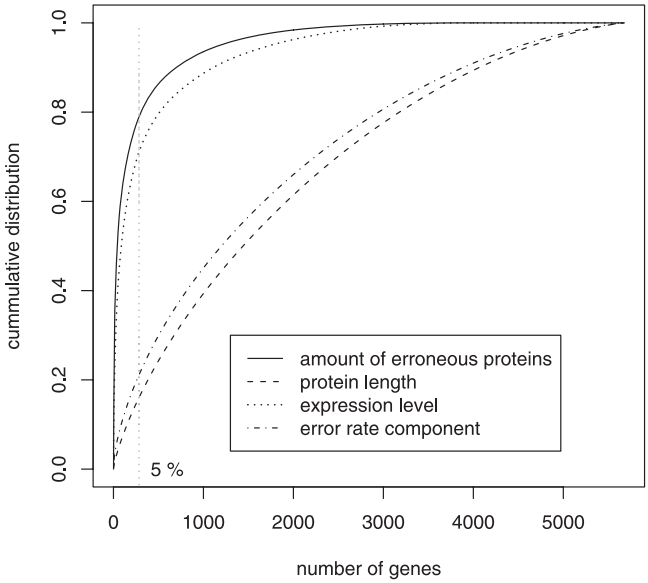

Figure 4. Relative Contribution of Each Gene to the Amount of Erroneous Proteins in Yeast if $u = 5 \times 10^{-4}$

The solid line shows the cumulative contribution of each gene to $\bar{x}$. A steep increase, as seen here, indicates that few genes are responsible for most of the abnormal proteins in a cell. We also plot the cumulative distribution for the three components of $n_i\bar{x}_i$: protein length, n_i (dashed line), number of functional proteins, y_i (dotted line), and expected amount of erroneous proteins to synthesize one error-free protein, $u_i/(1-u_i) = (1-(1-u)^{n_i})/(1-u)^{n_i}$ (dash-dotted line). The curvature of the solid line is similar to the curvature of the dotted line. Hence, most of the abnormal proteins stem from few genes because these genes are also expressed at a very high level.

doi:10.1371/journal.pcbi.0030203.g004

increases suddenly as u increases beyond 5×10^{-4}. For high values of u, phenotypic mutations are so frequent that it becomes essentially impossible to synthesize large proteins accurately. These large proteins dominate the cost of phenotypic errors. This can also be seen in the change of the weighted average of the expression level. For low mutation rates, the amount of erroneous proteins is dominated by highly expressed genes with a weighted average expression level of 2.67×10^5 proteins per cell. This value begins to decrease at $u = 5 \times 10^{-4}$ because large proteins, instead of highly expressed proteins, begin to increasingly contribute to the amount of abnormal proteins in the cell.

For $u = 5 \times 10^{-4}$, the bulk of the amount of erroneous proteins comes from proteins that are noticeably larger than the average protein and are highly expressed. How much these proteins contribute to $\bar{x}$ compared with the rest of the genome can be seen in Figure 4. The solid line shows $\sum_{i=1}^{k} n_i\bar{x}_i/\bar{x}$ for $k = 1, 2, \ldots, 5675$, that is, the cumulative contribution of each gene to $\bar{x}$. Genes are sorted decreasingly by their contribution to $\bar{x}$. It is obvious that only few genes contribute to most of the abnormal proteins in the cell. In fact, 5% (10%) of the genes contribute to 78.6% (87.5%) of the abnormal proteins in a yeast cell. The average length of these proteins is 927 (818) which confirms the conclusion from above that genes that contribute most to the amount of abnormal proteins are much larger than the average gene.

From Equation 7 we know that $n_i\bar{x}_i = n_i y_i(1-(1-u)^{n_i})/(1-u)^{n_i}$ and can distinguish three components: (i) the protein length, n_i, (ii) the expression level, y_i, and (iii) the expected number of erroneous proteins that have to be synthesized to get one error-free protein, $u_i/(1-u_i)$, with $u_i = 1-(1-u)^{n_i}$. Which of these components is primarily responsible for the fact that only few genes contribute to most of the abnormal proteins in a cell? To answer this question, we can compare the dashed, dotted, and dash-dotted lines in Figure 4 which show how unevenly genes contribute to each of the three components (for $u = 5 \times 10^{-4}$). For example, the dotted line shows $\sum_{i=1}^{k} y_i / \sum_{i=1}^{5675} y_i$ for $k = 1, 2, \ldots, 5675$. From the three components, only the expression level (dotted line) shows a curvature similar to the solid line. Hence, the fact that few genes contribute to most of the abnormal proteins in a cell is due to differences in expression levels rather than differences in protein length.

Adaptation of Highly Expressed Proteins

If most of the abnormal proteins in a cell are synthesized by few, highly expressed proteins, the cell could reduce the cost of phenotypic mutation rates considerably by decreasing the phenotpyic mutation rate for these few genes. In fact, highly expressed genes are special in many ways. They use preferred codons more frequently than "normal" genes [28] and evolve more slowly [29]. The usage of preferred codons conveys several advantages, among them is a more efficient [30–32] and accurate [33–35] translation. As argued by Drummond et al. [29], the slow rate of evolution of highly expressed genes might be the result of selection for translational robustness, that is, the ability of proteins to work properly despite amino acid substitutions [15,29]. The effect of preferred codon usage and translational robustness are conceptually very different. The usage of preferred codons reduces the phenotypic

mutation rate by reducing the amino acid substitution rate, while an increase in translational robustness reduces the phenotypic mutation rate by improving a protein's ability to withstand the effect of amino acid substitutions. Let us first consider translational robustness.

Selection for translational robustness

Let u_{aa} denote the amino acid substitution rate. Together with the robustness of a protein against amino acid substitutions it determines the protein's phenotypic mutation rate u_i. According to Bloom et al. [36], the probability that a protein retains its wild-type structure after m amino acid substitutions is given by

$$P_f(m) \propto v^m, \tag{12}$$

where v denotes the average neutrality to amino acid substitutions (m-neutrality), that is, the average probability that a protein will be unaffected by ("neutral" to) an additional amino acid substitution. In the following we make the conservative assumption that a protein is functional if it is able to fold into its wild-type structure and that the wild-type sequence always folds into its wild-type structure, i.e., that $P_f(0) = 1$. Consequently, $P_f(m) = v^m$ is the probability that a protein exposed to m amino acid substitutions is functional.

A protein contains m amino acid substitutions with probability

$$p(m) = \binom{n}{m}(1-u_{\mathrm{aa}})^{n-m}u_{\mathrm{aa}}^m \tag{13}$$

and is therefore functional with probability

$$p_{F,i} = \sum_{m=0}^{n_i} p(m)P_f(m) = \sum_{m=0}^{n_i}\binom{n_i}{m}(1-u_{\mathrm{aa}})^{n_i-m}u_{\mathrm{aa}}^m v_i^m. \tag{14}$$

The probability to synthesize a nonfunctional protein from gene i is given by

$$u_i = 1 - p_{F,i}. \tag{15}$$

For protein i, the average number of nonfunctional proteins is given by $\bar{x}_i = y_i u_i/(1-u_i) = y_i(p_{F,i}^{-1} - 1)$, and the overall average amount of nonfunctional proteins in a cell is given by

$$\begin{aligned}\bar{x} &= \sum_i n_i\bar{x}_i = \sum_i n_i y_i(p_{F,i}^{-1} - 1)\\ &= \sum_i n_i y_i\left(\left[\sum_{m=0}^{n_i}\binom{n_i}{m}(1-u_{\mathrm{aa}})^{n_i-m}(u_{\mathrm{aa}}v_i)^m\right]^{-1} - 1\right).\end{aligned} \tag{16}$$

In comparison with Equation 7, we note that $(1-u)^{-n_i}$, the term that is responsible for the rapid, nonlinear increase of $\bar{x}$, has been replaced by a sum over the number of amino acid substitutions. Since n_i is usually much larger than the number of amino acid substitutions, we have $n_i \approx n_i - \mathrm{m}$ for relevant m values and can approximate the sum very accurately by

$$(1-u_{\mathrm{aa}})^{n_i}\sum_{m=0}^{n_i}\binom{n_i}{m}(u_{\mathrm{aa}}v_i)^m. \tag{17}$$

We see that the term that caused the dramatic increase in the previous model also appears in this model, which considers translational robustness. The phenotypic mutation rate u is replaced by u_{aa} and the term $(1-u)^{n_i}$ multiplied by a factor that depends on the protein's m-neutrality v_i. Analogous to our previous observations, we can expect a nonlinear increase of $\bar{x}$ for $u_{aa} > 5\times 10^{-4}$.

In theory, but not in practice, it is possible to reduce the phenotypic mutation rate to zero by increasing v_i to 1 for all proteins. In practice, an upper limit for v_i is given by the function and stability of the protein. In our simulations, this upper limit is set by a prior distribution for the values of v_i. Very high values for v_i will be possible but unlikely. Given this (soft) upper limit for m-neutralities, we can, for a given amino acid substitution rate, ask which proteins will be selected for translational robustness (large m-neutralities) and what amount of abnormal proteins can be expected.

We present simulations in which m-neutralities are drawn from a beta distribution, $B(v|a,b) \propto v^{a-1}(1-v)^{b-1}$ with $a = 16.95$ and $b = 20.72$, which has variance 0.0064 and mean 0.45, reflecting the mean and variance of m-neutralities of seven proteins estimated by Bloom et al. [36]. We want to emphasis that the quantitative results, in particular the relative changes as a function of u_{aa}, are not affected when other (reasonable) prior distributions are used. We obtained similar results for a beta distribution with mean 0.5 and variance 0.02 and for corresponding normal distributions (truncated to the interval [0, 1]). Even though the absolute value of $\bar{x}$ is smaller for prior distributions that allow larger m-neutralities, the relative changes remain the same.

Mutations generate m-neutralities v_i for each protein from the prior distribution. Selection determines if an m-neutrality reaches fixation and subsequently the eventual distribution of v_i after many generations of selection. By how much selection has caused a protein's m-neutrality to deviate from the prior distribution can be expressed in terms of the log likelihood (LL) of the v_i values under the prior distribution, $\log(B(v_i|a,b)) \propto (a-1)\log(v_i) + (b-1)\log(1-v_i)$. If selection leads to a significant increase of a protein's m-neutrality, then the LL of this m-neutrality will be very low. The average LL for a cell's proteins is given by

$$\log\left(\prod_{i=1}^{K} B(v_i|a,b)\right)^{1/K} = \frac{1}{K}\sum_{i=1}^{K}\log[B(v_i|a,b)] \tag{18}$$

and can be used to quantify the extent of selection for translational robustness that the proteins of a cell were exposed to. As we will see, the LL of the m-neutralities (after selection) decreases substantially for $u_{\mathrm{aa}} > 5\times 10^{-4}$ and indicates the intensified selection for translational robustness.

We use a drift-selection model based on the fixation probability in a Moran process to determine the v_i values at mutation–selection balance (the post-selection v_i distribution). For a given amino acid substitution rate, we initialize v_i for all proteins by setting it equal to the mean of the prior distribution and calculate $\bar{x}$. For every protein (gene) i, we draw a new v_i from the prior distribution and calculate the new $\bar{x}$ that reflects this change, $\bar{x}_{\mathrm{new}} = \bar{x} + n_i y_i(p_{F,i,\mathrm{new}}^{-1} - p_{F,i}^{-1})$. For computational reasons and since the binomial distribution has most of its weight at small values of m, we truncate the summation over m in Equation 14 to the smallest integer larger than $n_i u_{aa} + 4\sqrt{n_i u_{\mathrm{aa}}(1-u_{\mathrm{aa}})}$ (i.e., m-values that are more than four standard deviations away from the mean are ignored; the probability of getting m's above this threshold is less than 2×10^{-3}). We accept the new v_i with probability $(1-1/r)/(1-1/r^N)$, which corresponds to the fixation probability in a

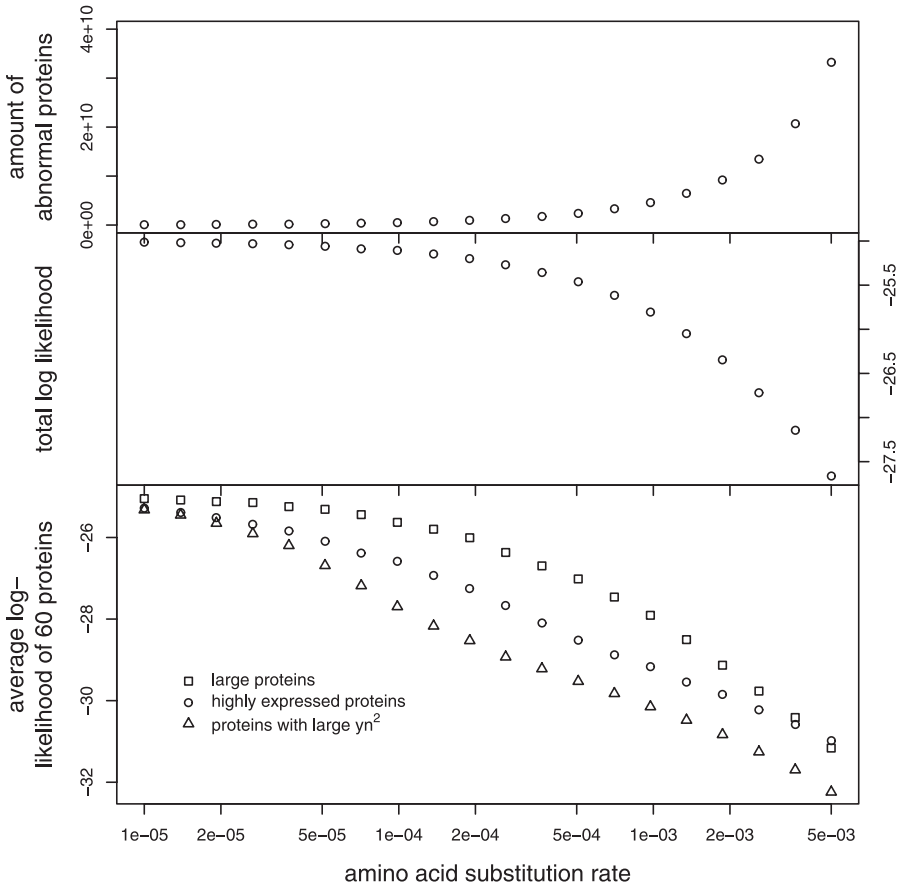

Figure 5. Selection for Translational Robustness as a Function of Amino Acid Substitution Rate

The top panel shows the average amount of abnormal proteins, $\bar{x}$, after 500,000 cycles of mutation and selection (see text for more detail). The mid-panel shows the average log-likelihood (LL) of the m-neutralities after selection. For $u \geq 1.92 \times 10^{-5}$, the average LLs are significantly lower than expected by chance. The 5 (0.1) quantile is given by −25.024 (−25.040). The total LL decreases noticeably for $u_{aa} > 5 \times 10^{-4}$. The lower panel shows the average LL of three groups of 100 proteins. We consider the 100 largest proteins, the 100 most highly expressed proteins, and the 100 proteins with the largest $y_i n_i^2$. For 100 proteins, the 5% (0.1%) quantile for the average LL is −25.117 (−25.215). The lower panel shows that the extent of selection for translational robustness increases nonlinearly for large proteins whereas it increases approximately linearly in the other two groups of proteins.

doi:10.1371/journal.pcbi.0030203.g005

Moran process of a mutant with relative fitness $r = f_{new}/f$ in a population of size N. We use $N = 10{,}000$ and $f(\bar{x}) = e^{-c\bar{x}}$, with $c = 10^{-9}$. Here, we cannot use the fitness function $1 - c\bar{x}$ as before, because it might lead to negative fitness values. In the previous section, we chose $1 - c\bar{x}$ because of its analytical tractability and its similarity to the cost of genotypic mutations (see Equation 1). We did not have to worry about negative values for $1 - c\bar{x}$ since u could evolve freely and selection caused u to converge to levels where $c\bar{x} \approx 5/N < 1$. In this section, u_{aa} is constant and the adaptation of individual proteins cannot reduce $\bar{x}$ arbitrarily (there are upper limits to v_i). Note that $1 - c\bar{x} \approx e^{-c\bar{x}}$ for small $c\bar{x}$. Hence, our results from the previous section, where we used $1 - c\bar{x}$ as cost function also hold for the cost function $e^{-c\bar{x}}$.

For each u_{aa}, we report the average of 20 simulations, which differ only with respect to the seed for the random number generator. For each simulation, we sequentially conducted 500,000 updates of each protein as described above. We then analyzed which proteins were selected for translational robustness by calculating the LL. We also analyzed to which extent $\bar{x}$ is reduced by increasing v_i and how an increase in u_{aa} affects the selection for translational robustness.

Figure 5 summarizes the results of our simulations. The top and the middle panels show $\bar{x}$ and the average LL (18) after selection as a function of u_{aa}. Similar to the previous section, we notice a dramatic increase of $\bar{x}$ for $u_{aa} > 5 \times 10^{-4}$. This is not surprising, considering the mentioned analytical similarities between Equations 7 and 16. For amino acid substitution rates above 5×10^{-4}, the cell has difficulties to prevent the increase of $\bar{x}$. This is also evidenced by the decline of the LL. The lower panel in Figure 5 shows the change in the average LL of three sets of 100 proteins. The three sets of proteins are given by the 100 proteins with the largest n_i, y_i, and $y_i n_i^2$, respectively. As expected, proteins with large $y_i n_i^2$ values are more effectively selected for large m-neutralities than large or highly expressed proteins and, accordingly, have the smallest LL. With increasing u_{aa}, large proteins contribute more to the amount of abnormal proteins and the corresponding LL decreases more rapidly than for the other two groups of proteins. In Figure 6 we show the m-neutralities of individual proteins for three amino acid substitution rates. It illustrates the intensified selection for large m-neutralities in large proteins.

The simulations conducted in this section implicitly assume that the population is homogeneous and that one mutant appears at a time and either goes extinct or gives rise to

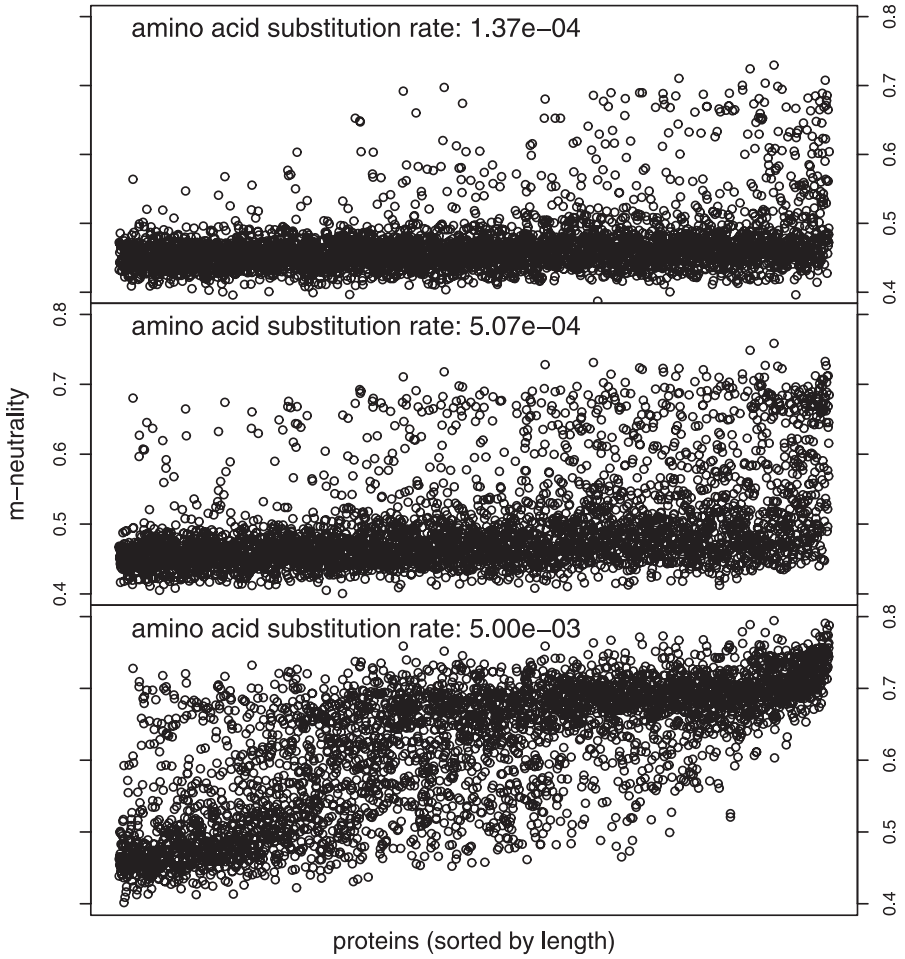

Figure 6. m-neutralities, v_i, After Selection for Three Different Amino Acid Substitution Rates u_{aa}

The data points are sorted according to the length of the proteins. The increased effectiveness of selection for higher m-neutralities in large proteins for high amino acid substitution rates is clearly visible.

doi:10.1371/journal.pcbi.0030203.g006

another homogeneous population. Hence, every organism in these simulations represents a homogeneous population of size N. This organism is of course also the MRCA of this population. We can compare its fitness, $f = e^{-c\bar{x}}$, with the fitness, $f = 1 - c\bar{x}$ ($\approx e^{-c\bar{x}}$ for $c\bar{x} \approx 0.005$), of the MRCA in our previous simulations to speculate on what would happen if we allowed u_{aa} to change here as well. Here, where u_{aa} was held constant, the fitness of the organism converged to fairly small values compared to the equilibrium values of $f \approx 0.995$ in our previous simulations which allowed changes in u. For $u_{aa} = 1 \times 10^{-5}$, where selection for higher m-neutralities is insignificant, $f = 0.938$, and f is much lower for larger phenotypic mutation rates (e.g., $f = 0.090$ for $u_{aa} = 5 \times 10^{-4}$). This suggests that if u_{aa} is able to evolve freely to equilibrium values of $f \approx 0.995$, then selection for translational robustness will be insignificant. Simulations in which we mutated u_{aa} as described in the previous section confirmed this expectation. No significantly elevated m-neutralities evolved (unpublished data).

This was not the case in simulations with few (e.g., ten) genes, where the LL of highly expressed proteins converged to significantly lower values. Apparently, if there are many genes and if u_{aa} is in mutation–selection balance, the m-neutralities of individual proteins do not contribute enough to allow selection for higher m-neutralities. But as we have seen above, if u_{aa} is constant, selection for larger m-neutralities can reduce $\bar{x}$ to some extent. Hence, if u_{aa} is above its mutation–selection-balance value, then significantly higher m-neutralities will evolve and decrease the phenotypic mutation rate by decreasing the effect of amino acid substitutions.

Selection for Preferred Codons

Besides increasing the translational robustness of certain proteins, a cell can also use preferred codons to decrease the phenotypic mutation rate u_i. This would actually decrease the amino acid substitution rate and is therefore conceptually different from translational robustness, which reduces the effect of amino acid substitutions but not their occurrence. Considering codon usage, the amino acid substitution rate u_{aa} has two components, a ribosomal component u_r and a codon-based component, u_c. We assume that $u_{aa} = u_r u_c$ for preferred codons and that $u_{aa} = u_r$ for nonpreferred codons. Preferred codons are more accurate than nonpreferred codons, hence, $u_c < 1$. In this section we ignore translational robustness, i.e., $u = u_{aa}$. A protein of length n_i that uses $\tilde{n}_i$ preferred codons synthesizes a functional protein with probability $(1 - u_r)^{n_i - \tilde{n}_i}(1 - u_r u_c)^{\tilde{n}_i}$. The average amount of abnormal proteins is given by

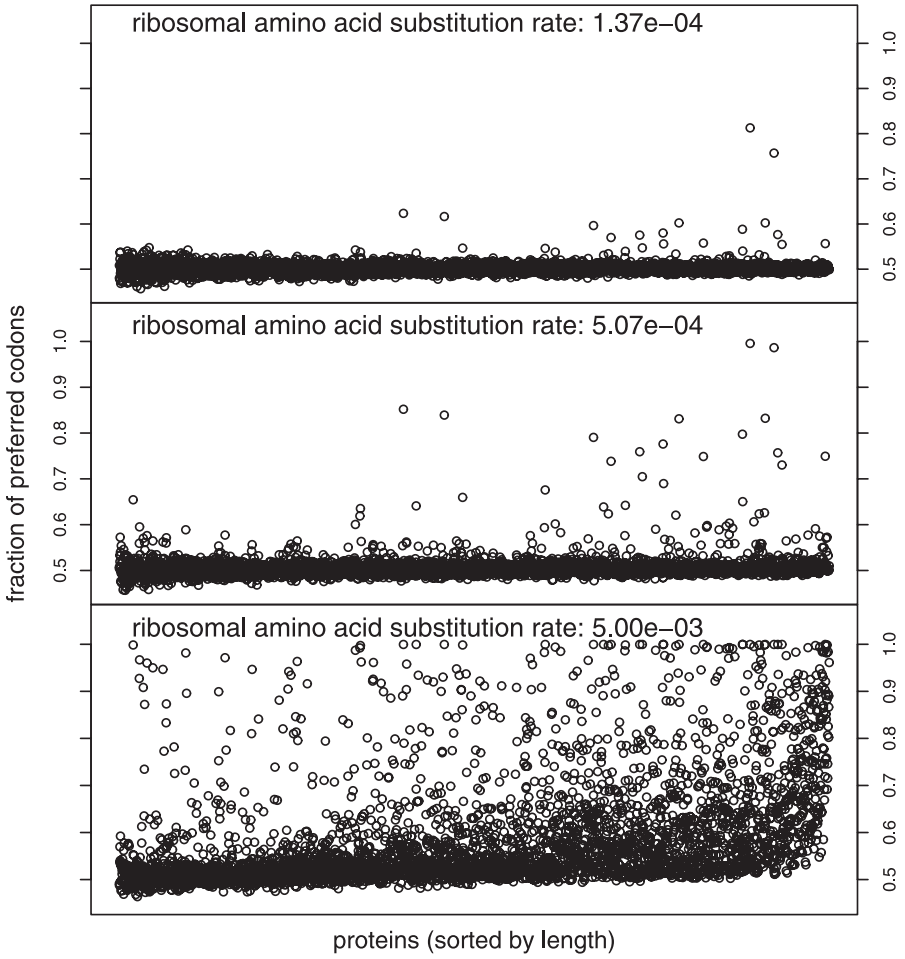

Figure 7. Fraction of Preferred Codons ($p_i = \tilde{n}_i/n_i$) After Selection for Three Different Ribosomal Amino Acid Substitution Rates u_r

The genes are sorted according to the length of the protein. For $u_r = 1.37 \times 10^{-4}$, selection introduces a codon bias only for few genes (compare with m-neutralities in Figure 6). An increase in the phenotypic mutation rate leads to more intensive selection for preferred codons in large genes than in small genes.

doi:10.1371/journal.pcbi.0030203.g007

$$\bar{x} = \sum_i n_i y_i [(1-u_r)^{-(n_i-\tilde{n}_i)}(1-u_r u_c)^{-\tilde{n}_i} - 1] \qquad (19)$$

$$= \sum_i n_i y_i [(1-u_r)^{-n_i}\left(\frac{1-u_r u_c}{1-u_r}\right)^{-\tilde{n}_i} - 1] \qquad (20)$$

Again, we notice similarities between Equations 7 and 19 and can expect a rapid increase of $\bar{x}$ for u_r increasing above 5×10^{-4}.

We conducted simulations analogous to those investigating the effect of translational robustness. We calculate $\bar{x}$ according to Equation 19 with $u_c = 0.1$. Each time we mutate the number of preferred codons, we increase $\tilde{n}_i$ by one with probability $(n_i - \tilde{n}_i)/n_i$ (the fraction of nonpreferred codons in the gene). We decrease $\tilde{n}_i$ by one with probability $\tilde{n}_i/n_i$. After changing $\tilde{n}_i$, we calculate $\bar{x}_{\text{new}}$ and accept the new $\tilde{n}_i$ as described in the section Selection for Translational Robustness. We report the average of 20 simulations for each u_r. Each simulation was terminated after 2×10^7 sequential mutations (not necessarily fixation) of $\tilde{n}_i$.

Figure 7 shows the results of our simulations. Analogous to Figure 6, we plot the equilibrium fraction of preferred codons, $p_i = \tilde{n}_i/n_i$, for each gene for three ribosomal amino acid substitution rates u_r. For $u = 1.37 \times 10^{-4}$, only few genes evolve a major codon bias of $p_i > 0.6$. The gene with the largest codon bias of about 0.8 encodes for the protein with the largest $y_i n_i^2$ and contributes 5.7% to the total amount of functional proteins and 7.9% to $\sum y_i n_i^2$.

For $u_r > 5 \times 10^{-4}$, large proteins begin to contribute more to the amount of abnormal proteins and selection increases the codon bias of large proteins. Similar to our observation in Figure 2, the codon bias cannot prevent the drastic increase of $\bar{x}$ for $u_r > 5 \times 10^{-4}$. As before, no significant codon bias evolved if we allowed u_r to change as well.

Comparing Figure 6 with Figure 7, we notice that selection for translational robustness results in a more distinct bias in v_i than what we observe for p_i after selection for preferred codons. In the next section we compare the two mechanisms to identify the source of this difference.

Similarities between Preferred Codons and Translational Robustness

In our simulations, the two mechanisms differ in the way $p_{F,i}$, the probability of synthesizing a functional protein, is calculated and in the way it is mutated.

For the translational robustness model, we calculate $p_{F,i}$ as

$$p_{F,i} = \sum_{m=0}^{n_i} \binom{n_i}{m} (1-u_{aa})^{n_i-m} u_{aa}^m v_i^m \approx (1-u_{aa})^{n_i} \sum_{m=0}^{n_i} \binom{n_i}{m} (u_{aa} v_i)^m. \quad (21)$$

The approximations are accurate as long as proteins with many amino acid substitutions are rare. For the preferred codon model, we have

$$\begin{aligned} p_{F,i} &= (1-u_r)^{n_i-\tilde{n}_i}(1-u_r u_c)^{\tilde{n}_i} \\ &= (1-u_r)^{n_i}\left[1+u_r\frac{1-u_c}{1-u_r}\right]^{n_i p_i} \\ &\approx (1-u_r)^{n_i}[1+u_r p_i(1-u_c)]^{n_i} \\ &= (1-u_r)^{n_i}\sum_{m=0}^{n_i}\binom{n_i}{m}[u_r p_i(1-u_c)]^m, \end{aligned} \quad (22)$$

where we used $1+u_r/(1-u_r) \approx 1+u_r$ and $[1+u_r(1-u_c)]^{p_i} \approx 1+u_r p_i(1-u_c)$, which are reasonable approximations if u_r is small. The analogy between Equation 21 and Equation 22 is obvious. In theory, the m-neutralities, v_i, can range from 0 to 1. In our simulations, for the chosen prior distribution, v_i's larger than 0.8 are rare. The m-neutralities v_i are analogous to the term $p_i(1 - u_c)$ in the preferred codon model. Since p_i can range from 0 to 1, the two mechanisms can reduce the amount of abnormal proteins equally well if $v_{max} = 1 - u_c$, where v_{max} denotes the upper limit for m-neutralities. In our simulations, we have $v_{max} \approx 0.8 < 1 - u_c = 0.9$. Hence, we would expect lower $\bar{x}$ values in the preferred codon model. This is not the case. For example, for $u_{aa} = 10^{-5}$, $\bar{x}$ converged to 6.4×10^7 in the translational robustness model, whereas $\bar{x}$ converged to 7.3×10^7 in the preferred codon model. Hence, we have to consider the way in which the $p_{F,i}$'s are mutated to understand this result.

In the translational robustness model, v_i is sampled from a prior distribution. The new v_i value is independent of the previous one. Hence, large changes of v_i and, consequently, of $p_{F,i}$ are possible. In the preferred codon model, the number of preferred codons can only change in increments of one, and corresponding changes of $p_{F,i}$ and $\bar{x}$ are small. The small changes in p_i and, therefore, in $p_{F,i}$ allow the evolution of noticeable codon biases only in genes that produce large amounts of abnormal proteins. In the translational robustness model, large changes in the $p_{F,i}$'s are possible, and they have a higher fixation probability.

Take, for example, the protein with the largest value of $y_i n_i^2$. It is 918 amino acids long, and a change of $\tilde{n}_i$ from 459 to 460 increases p_i from 0.5 to 0.501 (by 0.2%). This small change reaches fixation only if the costs of abnormal proteins from this gene are very large. Changes larger than this are frequent in the translational robustness model and have a higher probability of fixation.

Discussion

A functional protein machinery, built from genetic information, is central to every living organism. Surprisingly, the decoding of genes into amino acid sequences is fairly inaccurate. Errors (phenotypic mutations) occur several orders of magnitude more frequently than during DNA replication. The frequency of errors depends on the codon and its context (see Table 1).

In this paper, we have explored the evolution of phenotypic mutation rates. In our model, a cell maintains protein synthesis until a certain number of functional proteins are present. Depending on the phenotypic mutation rate, u, a certain number of amino acids, x, are "wasted" in erroneous proteins and reduce the fitness of the organism by $\eta(x)$. For simplicity, we used a linear cost function $\eta(x) = cx$. With genomic and proteomic data from *S. cerevisiae* [24,25], we discover (a) an effective upper bound for the phenotypic mutation rate, (b) that most of the abnormal proteins stem from genes that are highly expressed and substantially larger than the average yeast protein, (c) that an average phenotypic mutation rate of $u = 5 \times 10^{-4}$ is at a value where x begins to increase dramatically as a function of u and large, lowly expressed genes begin to contribute substantially to the amount of abnormal proteins, and (d) that an increased codon bias or translational robustness in highly expressed genes can reduce the amount of abnormal proteins but cannot stop the dramatic increase for amino acid substitution rates above 5×10^{-4}.

To what extent do our results depend on the assumption that $\eta(x)$ is linear and that gene expression is maintained until a certain number of functional proteins are present? Dekel and Alon [37] found a convex increase of the cost of protein synthesis with the amount of proteins synthesized. Considering this and that aggregates of misfolded proteins are, in a concentration-dependent way, toxic to cells [27,38], we can expect the cost of erroneous proteins to increase faster than linear with the amount of erroneous proteins produced. A nonlinear $\eta(x)$, however, would only affect the position of the upper bound for u, which we observed in our simulations (see Figure 1). For a nonlinear cost function, we would expect this upper bound to be lower than what we have observed here, because of the nonlinear increase of x on top of the nonlinear increase of the costs of x. Results (b)–(d) are not affected by the shape of the cost function.

Let us now consider our assumption about the regulation of gene expression. The largest protein in the yeast genome is Mdn1p, a dynein-related AAA-type ATPase [24,39]. It is 4,910 amino acids long. For $u = 5 \times 10^{-4}$, only 12.8% of the synthesized proteins are error-free. To get the required number of 0.538×10^3 error-free proteins [25], the cell has to synthesize 6×10^3 proteins. This is not a tremendous burden considering that about $46{,}600 \times 10^3$ functional proteins are synthesized in total. However, this number increases rapidly if u increases. Doubling or quadrupling u would require the synthesis of 72.7×10^3 or $10{,}000 \times 10^3$ proteins, respectively. It is unrealistic to assume that a cell will synthesize 10^7 proteins to get 538 functional ones. But we can consider this rapid increase as an indication for the inability of the cell to synthesize this protein and would have to rephrase result (c) to account for our assumption about gene expression: (c′) a phenotypic mutation rate of $u = 5 \times 10^{-4}$ is at a value where it is still feasible to synthesize large proteins. Higher phenotypic mutation rates would make it impossible to synthesize large proteins.

Interestingly, if one considers the ability of the cell to synthesize a certain number of functional proteins after a certain number of synthesis attempts, an upper bound for u is also encountered. In this situation, however, this upper bound is not due to the increase in abnormal proteins and the associated cost but due to the inability of the cell to synthesize enough functional proteins. In such a situation the

cost of abnormal proteins is largely irrelevant and the upper bound for u primarily a result of the protein-length distribution and not of the cost of abnormal proteins. Furthermore, in such a situation there is little selection pressure to reduce the phenotypic mutation rate much below this upper bound [16].

If the synthesis of large proteins is such a problem, why does the cell not synthesize many smaller proteins and assemble them after successful production? An intermediate check for proper folding (which equals proper function for most amino acid substitutions) would prevent the incorporation of nonfunctional subunits and reduce the probability of assembling a nonfunctional complex. In yeast, proteins with a length of about 1,000 amino acids are quite common. This suggests that the complexation of proteins much smaller than 1,000 amino acids constitutes a considerable challenge. For so many large proteins, it might be impossible to get the same biological function from a complex of smaller proteins. According to our model, an upper bound of 1,000 for the yeast protein length does not reduce the drastic increase by much. If we calculate $\bar{x}$ after removing all proteins from the dataset that are larger than 1,000 amino acids, we can still observe a rapid increase in $\bar{x}$ at $u = 5 \times 10^{-4}$; doubling (quadrupling) u would lead to a 2.4 (7.3)-fold increase in the amount of abnormal proteins. Therefore, partitioning extremely large proteins into protein complexes is not sufficient to avoid the negative effects of an increasing phenotypic mutation rate.

Instead of complexing large proteins, evolution could reduce the phenotypic mutation rate of individual proteins. The phenotypic mutation rate of individual proteins could be reduced by using preferred codons [33–35] or by increasing the translational robustness of proteins [15,29,40]. Our analysis shows that these two mechanisms have nearly the same potential to minimize $\bar{x}$ if u_c is sufficiently small (i.e., if preferred codons are sufficiently more accurate than nonpreferred codons). One big difference between preferred codons and translational robustness is the way in which the trait is mutated. For preferred codon usage, it seems reasonable to assume that the number of preferred codons changes in increments of one, which leads to very small changes in the amount of abnormal proteins.

Considering translational robustness, little is known about how mutations change the translational robustness of a protein. In our simulations, we mutate the translational robustness of a protein by sampling it from a prior distribution, which allows for large changes. Alternatively, one can use models that allow only small changes in a protein's translational robustness. More empirical data on the translational robustness spectrum of proteins is necessary to develop a satisfying model.

The effect of incremental changes of the number of preferred codons on the amount of abnormal proteins is fairly small. An increase in the number of preferred codons by one increases the probability of synthesizing a functional protein only by a factor of $(1 - u_r u_c)/(1 - u_r)$. For $u = 5 \times 10^{-4}$ and $u_c = 0.1$, this equals 1.00045. Since only few genes contribute much to the amount and number of abnormal proteins, this will lead to very small changes of $\bar{x}$ for most proteins.

As mentioned previously, preferred codons are also able to increase the rate of translation. Selection for faster translation (or higher expression level) could be responsible for the observed codon biases. Since the time it takes to synthesize y_i functional proteins is proportional to $y_i n_i$ and the amount of erroneous proteins is approximately proportional to $y_i n_i^2$, it is possible to distinguish between the two sources of codon bias by comparing the observed codon bias in yeast with the predicted codon bias if selective forces were proportional to $y_i n_i$ or $y_i n_i^2$.

Further, a refined version of our preferred codon model that considers the genetic code and the actual amino acid sequence of each yeast protein could be used to estimate the cost of abnormal proteins and the amino acid substitution rate. For a given amino acid substitution rate, u_r, an increase of the cost of abnormal proteins, c, increases the extent of codon bias but does not affect its distribution with respect to the protein length (the points in the top panel of Figure 7 would all move upward by an amount that is independent of n_i since $n_i \bar{x}_i$ remains unchanged for constant u_r). For given c, an increase of u_r changes the extent of codon bias as well as the codon bias distribution with respect to the protein length (as seen in Figure 7, if u_r increases, the codon bias of large proteins changes to a greater extent than the codon bias of small proteins since $\bar{x}_i$ will increase more for genes with large n_i). Hence, by choosing different values for c and u_r and by comparing the resulting extent and distribution (with respect to n_i) of codon biases with the extent and distribution of codon bias found in yeast, one can estimate the two parameters.

To experimentally measure the rate of amino acid substitutions during protein synthesis is notoriously difficult. Abnormal proteins are difficult to detect and usually degraded within minutes [27]. Experiments are usually limited to measuring the rate of specific substitutions at specific sites (see Table 1). One exception is work by Ellis and Gallant [4], who measured the rate of substitution of charged amino acids by uncharged amino acids. For many proteins such substitutions are detectable as satellite spots after 2-D gel electrophoresis. However, their method might fail to detect rapidly degraded abnormal proteins and is dependent on the number of codons at which charge substitutions can occur [4].

It would be highly desirable to be able to calculate the actual frequency of phenotypic mutations, that is, the frequency of deleterious amino acid substitutions during protein synthesis as opposed to the frequency of all (detrimental or not) amino acid substitutions. We can use our model together with data on the fraction of proteins that are abnormal and degraded rapidly [41,42] to calculate this. Schubert et al. [41] and Princiotta et al. [42] measured that in human cells about 33% and 25%, respectively, of newly synthesized proteins are rapidly degraded. The proteins are degraded mainly because of their inability to achieve a functional state [27]. Since these are values for human cells and might also include proteins that could not achieve a functional state despite error-free protein synthesis, we will use 15%–35% as the range for the fraction of proteins that are nonfunctional due to phenotypic mutations. In our model, y and $\bar{x}$ give the amount of functional and nonfunctional proteins synthesized, respectively. Hence the fraction of nonfunctional proteins synthesized due to phenotypic errors is given by $\bar{x}/(y + \bar{x})$. According to our model (Equation 7) and the data from yeast (see Materials and Methods), $2.4 \times$

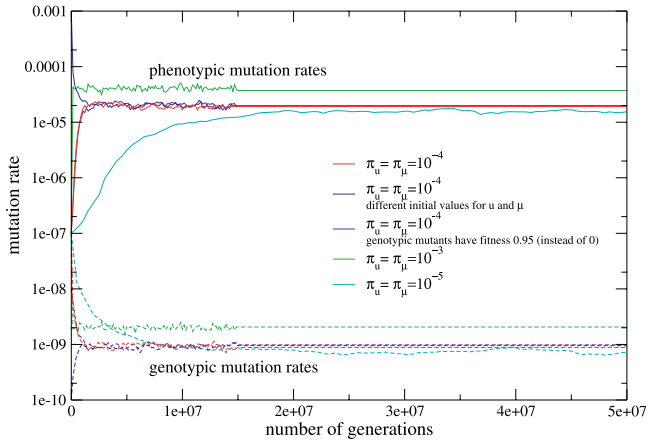

Figure 8. Evolution of Genotypic and Phenotypic Mutation Rates with Different Simulation Parameters (see the section Effect of Initial Values and Parameters on the Simulation Results)

For all five parameter combinations, we use $c = 10^{-11}$. Most simulations reach equilibrium within 1.5×10^7 generations. For $\pi_u = \pi_\mu = 10^{-5}$, it was necessary to extend the simulations to 5×10^7 generations. For simulations that end after 1.5×10^7 generations, we plot straight lines thereafter. These lines serve as visual cues and are at the level of the last value.

doi:10.1371/journal.pcbi.0030203.g008

10^{-4} to 6.1×10^{-4} deleterious amino acid substitutions per codon would result in the synthesis of 15% to 35% nonfunctional proteins. Better estimates of the fraction of abnormal proteins in yeast would allow a narrowing of the calculated range.

Mutation Rates at Equilibrium

Here, we derive our main analytical results on the magnitude of the genotypic and phenotypic mutation rates stated in the section The Model. We start by recalling Haldane's principle for an asexually reproducing population. This population is assumed to be sufficiently large so that random genetic drift can be ignored. The only evolutionary forces considered are selection and mutation. We assume that there is an optimal type (wild type) in this population. Its fitness is denoted by W_0, the rate at which mutations to other types occurs is denoted by U, and back mutations are ignored. Then the mean fitness of the population at mutation–selection balance is given by $\bar{W} = (1 - U)W_0$. This is obtained immediately from the recursion relation $p_0' = (1 - U)W_0 p_0/\bar{W}$ for the frequency p_0 of the optimal type. The important, but simple point, first made by Haldane, is that the mean fitness is independent of the fitnesses of the deleterious types ([43], pp. 106–107).

This principle can be generalized to a large class of mutation patterns among possible types, and even to a continuum of possible types. It then states that in mutation–selection balance mean fitness $\bar{W}$ satisfies $(1 - U)W_0 < \bar{W} < W_0$, where every type in the population is assumed to have the same mutation rate U. In addition, $\bar{W}$ becomes asymptotically equal to $(1 - U)W_0$ if the mutation rate U becomes sufficiently small. Detailed formulations as well as proofs can be found in ([43], pp. 127, 143–148). Again, the equilibrium mean fitness is, to first order in U, independent of the precise mutation pattern and of the fitnesses of the deleterious types.

Now we derive approximations for $\hat{\mu}$ and $\hat{u}$ in our model. We assume that the cost function η is linear, i.e., $\bar{\eta}(\bar{x}) = 1 - c\bar{x}$. Because of its complexity, we need a simplified model to make analytical progress. We identify all cells that have the same pair of mutation rates, (μ,u), and assign to them the average fitness $\bar{f}(\mu, u)$ (see Equation 1) of a population of cells with these mutation rates. For given μ and applying Haldane's generalized principle to the trait "phenotypic mutation rate," we get

$$\bar{f}(\mu, u) = \bar{W} \approx \bar{f}(\mu, 0)(1 - \pi_u). \tag{23}$$

Rearrangement and use of Equation 8 yields the following approximation for the evolved phenotypic mutation rate at equilibrium:

$$\hat{u} \approx \pi_u \frac{1}{c \sum_i n_i^2 y_i}. \tag{24}$$

For the evolved genotypic mutation rate, we already have derived the approximation (Equation 4). The general theory [43], as well as numerical results (unpublished data), show that the above approximations for $\hat{\mu}$ and $\hat{u}$ are slight overestimates of the true values. Taking the ratio of Equation 24 and Equation 4, we obtain Equation 9.

Effect of Initial Values and Parameters on the Simulation Results

To show the robustness of our results with respect to the initial conditions and the parameters, we conducted additional simulations analogous to the simulations presented in Figure 1. For Figure 1, we used $\pi_u = \pi_\mu = 10^{-4}$ and 10^{-7} as initial values of u and μ; genotypic mutations were lethal. The blue and violet lines in Figure 8 show that the initial values for u and μ and the fitness of the genotypic mutant do not change the equilibrium mutation rates at mutation–selection balance. The genotypic and phenotypic mutation rates will converge to the same equilibrium mutation rates as long as (a) the initial value for u is low enough so that $f = 1 - \eta(\bar{x}) > 0$, and (b) the initial value for μ is low enough (or genotypic mutations deleterious enough) so that a fixation of genotypic mutants does not occur.

We conducted simulations with different values for π_u and π_μ. The green and cyan lines in Figure 8 show the evolution of u and μ for $\pi_u = \pi_\mu = 10^{-3}$ and $\pi_u = \pi_\mu = 10^{-5}$, respectively. As expected, higher (lower) π_u and π_μ lead to faster (slower) evolution of u and μ and to increased (decreased) equilibrium values. The magnitude of this change is smaller than predicted by theory, e.g., Equation 9. This can be attributed to the finite population size, $N = 10^4$. In finite populations, selection is inefficient for costs (C_u,C_μ) below a certain threshold. Note that from Equations 23 and 2 we have $\bar{f}(\mu, u) \approx \bar{f}(\mu, 0)(1 - \pi_u)$ and $\bar{f}(\mu, u) \approx \bar{f}(\mu, 0)(1 - C_u)$, respectively.

Acknowledgments

We thank two anonymous reviewers for their detailed and constructive comments.

Author contributions. MW, RB, and MAN conceived and designed the experiments and wrote the paper. MW performed the experiments. MW and RB analyzed the data.

Funding. MW was supported by a Merck-Wiley fellowship. Support from the NSF/NIH joint program in mathematical biology (NIH grant

r01gm078986) is gratefully acknowledged. The Program for Evolutionary Dynamics at Harvard University is sponsored by J. Epstein.

Competing interests. The authors have declared that no competing interests exist.

References

1. Blank A, Drake JW, Charlesworth B, Charlesworth D, Crow JF (1998) Rates of spontaneous mutation. Genetics 148: 1667–1686.
2. Blank A, Gallant JA, Burgess RR, Loeb LA (1986) An rna polymerase mutant with reduced accuracy of chain elongation. Biochemistry 25: 5920–5928.
3. Kurland CG (1992) Translational accuracy and the fitness of bacteria. Annu Rev Genet 26: 29–50.
4. Ellis N, Gallant J (1982) An estimate of the global error frequency in translation. Mol Gen Genet 188: 169–172.
5. Parker J (1989) Errors and alternatives in reading the universal genetic code. Microbiol Rev 53: 273–298.
6. Eigen M, Schuster P (1977) The hypercycle. A principle of natural self-organization. Part A: Emergence of the hypercycle. Naturwissenschaften 64: 541–565.
7. Eigen M, Mccaskill J, Schuster P (1988) Molecular quasi-species. J Phys Chem 92: 6881–6891.
8. Nowak M, Schuster P (1989) Error thresholds of replication in finite populations, mutation frequencies, and the onset of muller's ratchet. J Theor Biol 137: 375–395.
9. Nowak M (2006) Evolutionary dynamics: Exploring the equations of life. Cambridge (Massachusetts): Belknap Press.
10. Taddei F, Radman M, Maynard-Smith J, Toupance B, Gouyon PH, et al. (1997) Role of mutator alleles in adaptive evolution. Nature 387: 700–702.
11. Johnson T (1999a) Beneficial mutations, hitchhiking and the evolution of mutation rates in sexual populations. Genetics 151: 1621–1631.
12. Johnson T (1999b) The approach to mutation–selection balance in an infinite asexual population, and the evolution of mutation rates. Proc Biol Sci 266: 2389–2397.
13. Sniegowski PD, Gerrish PJ, Johnson T, Shaver A (2000) The evolution of mutation rates: Separating causes from consequences. Bioessays 22: 1057–1066.
14. André JB, Godelle B (2006) The evolution of mutation rate in finite asexual populations. Genetics 172: 611–626.
15. Wilke CO, Drummond DA (2006) Population genetics of translational robustness. Genetics 173: 473–481.
16. Bürger R, Willensdorfer M, Nowak MA (2006) Why are phenotypic mutation rates much higher than genotypic mutation rates? Genetics 172: 197–206.
17. Kurland C, Hughes D, Ehrenberg M (1996) *Escherichia coli* and *Salmonella*: Cellular and molecular biology. 2nd edition. Washington (D.C.): American Society for Microbiology. Chapter, Limitations of translational accuracy. pp. 979–1004.
18. Mikkola R, Kurland CG (1992a) Is there a unique ribosome phenotype for naturally occurring *Escherichia coli*? Biochimie 73: 1061–1066.
19. Mikkola R, Kurland CG (1992b) Selection of laboratory wild-type phenotype from natural isolates of *Escherichia coli* in chemostats. Mol Biol Evol 9: 394–402.
20. Hartl DL, Dykhuizen DE (1984) The population genetics of *Escherichia coli*. Annu Rev Genet 18: 31–68.
21. Hartl DL, Clark AG (1997) Principles of population genetics. 3rd edition. Sunderland (Massachusetts): Sinauer Associates.
22. Saunders CT, Baker D (2002) Evaluation of structural and evolutionary contributions to deleterious mutation prediction. J Mol Biol 322: 891–901.
23. Shaw RJ, Bonawitz ND, Reines D (2002) Use of an in vivo reporter assay to test for transcriptional and translational fidelity in yeast. J Biol Chem 277: 24420–24426.
24. Hong E, Balakrishnan R, Christie K, Costanzo M, Dwight S, et al. (2006) *Saccharomyces* genome database. Available: http://www.yeastgenome.org. Accessed 18 October 2007.
25. Ghaemmaghami S, Huh WK, Bower K, Howson RW, Belle A, et al. (2003) Global analysis of protein expression in yeast. Nature 425: 737–741.
26. Bremer H, Dennis P (1996) *Escherichia coli* and *Salmonella*: Cellular and molecular biology. Washington (D.C.): American Society for Microbiology. Chapter, Modulation of chemical composition and other parameters of the cell by growth rate. pp. 1553–1569.
27. Goldberg AL (2003) Protein degradation and protection against misfolded or damaged proteins. Nature 426: 895–899.
28. Gouy M, Gautier C (1982) Codon usage in bacteria: Correlation with gene expressivity. Nucleic Acids Res 10: 7055–7074.
29. Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH (2005) Why highly expressed proteins evolve slowly. Proc Natl Acad Sci U S A 102: 14338–14343.
30. Kurland CG (1991) Codon bias and gene expression. FEBS Lett 285: 165–169.
31. Xia X (1998) How optimized is the translational machinery in *Escherichia coli*, *Salmonella typhimurium* and *Saccharomyces cerevisiae*? Genetics 149: 37–44.
32. Gilchrist MA, Wagner A (2006) A model of protein translation including codon bias, nonsense errors, and ribosome recycling. J Theor Biol 239: 417–434.
33. Akashi H (1994) Synonymous codon usage in *Drosophila melanogaster*: Natural selection and translational accuracy. Genetics 136: 927–935.
34. Eyre-Walker A (1996) Synonymous codon bias is related to gene length in *Escherichia coli*: Selection for translational accuracy? Mol Biol Evol 13: 864–872.
35. Marais G, Duret L (2001) Synonymous codon usage, accuracy of translation, and gene length in caenorhabditis elegans. J Mol Evol 52: 275–280.
36. Bloom JD, Silberg JJ, Wilke CO, Drummond DA, Adami C, et al. (2005) Thermodynamic prediction of protein neutrality. Proc Natl Acad Sci U S A 102: 606–611.
37. Dekel E, Alon U (2005) Optimality and evolutionary tuning of the expression level of a protein. Nature 436: 588–592.
38. Bucciantini M, Giannoni E, Chiti F, Baroni F, Formigli L, et al. (2002) Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. Nature 416: 507–511.
39. Garbarino JE, Gibbons IR (2002) Expression and genomic analysis of midasin, a novel and highly conserved aaa protein distantly related to dynein. BMC Genomics 3: 18.
40. Wilke CO, Bloom JD, Drummond DA, Raval A (2005) Predicting the tolerance of proteins to random amino acid substitution. Biophys J 89: 3714–3720.
41. Schubert U, Ant'on LC, Gibbs J, Norbury CC, Yewdell JW, et al. (2000) Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. Nature 404: 770–774.
42. Princiotta MF, Finzi D, Qian SB, Gibbs J, Schuchmann S, et al. (2003) Quantitating protein synthesis, degradation, and endogenous antigen processing. Immunity 18: 343–354.
43. Bürger R (2000) The mathematical theory of selection, recombination, and mutation. Chichester (United Kingdom): John Wiley.
44. Parker J, Johnston TC, Borgia PT, Holtz G, Remaut E, et al. (1983) Codon usage and mistranslation. In vivo basal level misreading of the ms2 coat protein message. J Biol Chem 258: 10007–10012.
45. Bouadloun F, Donner D, Kurland CG (1983) Codon-specific missense errors in vivo. EMBO J 2: 1351–1356.
46. Edelmann P, Gallant J (1977) Mistranslation in *E. coli*. Cell 10: 131–137.
47. Toth MJ, Murgola EJ, Schimmel P (1988) Evidence for a unique first position codon-anticodon mismatch in vivo. J Mol Biol 201: 451–454.